Biometrics 71, 267–273
March 2015

DOI: 10.1111/biom.12228

READER REACTION

Reader Reaction to "A Robust Method for Estimating Optimal Treatment Regimes" by Zhang et al. (2012)

Jeremy M. G. Taylor,* Wenting Cheng, and Jared C. Foster

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
**email:* jmgt@umich.edu

Summary. A recent article (Zhang et al., 2012, *Biometrics* **168**, 1010–1018) compares regression based and inverse probability based methods of estimating an optimal treatment regime and shows for a small number of covariates that inverse probability weighted methods are more robust to model misspecification than regression methods. We demonstrate that using models that fit the data better reduces the concern about non-robustness for the regression methods. We extend the simulation study of Zhang et al. (2012, *Biometrics* **168**, 1010–1018), also considering the situation of a larger number of covariates, and show that incorporating random forests into both regression and inverse probability weighted based methods improves their properties.

Key words: Optimal treatment regime; Random forests.

1. Introduction

In an excellent article (Zhang et al., 2012), on estimating an optimal treatment regime, the authors consider the following situation: n subjects in a study, who are either in the treatment ($A = 1$) or the control ($A = 0$) group. Each subject has p baseline covariates $\boldsymbol{X} = (X_1, ..., X_p)$ and higher values of the continuous outcome measure (Y) are better. A treatment regime $g(\boldsymbol{X})$ is a function from $\boldsymbol{X}$ to $\{0, 1\}$, such that patients should receive $A = 1$ if $g(\boldsymbol{X}) = 1$ and $A = 0$ if $g(\boldsymbol{X}) = 0$. The value of $g(\boldsymbol{X})$ is determined by whether $\eta_0 + \sum_{j=1}^{p} \eta_j X_j$ is positive or not. The goal is to find the optimal treatment regime. Both a randomized trial and an observational study setting were considered. The authors develop and compare different approaches. One is a regression approach (RG), which requires a model for $\mu(A, \boldsymbol{X}) = E(Y|A, \boldsymbol{X})$. Other approaches are based on inverse probability weighted estimators (IPWE). The standard IPWE does not require a model for $\mu(A, \boldsymbol{X})$, but does requires a model for $P(A = 1|\boldsymbol{X})$. The authors extend the IPWE to an augmented inverse probability weighted estimator (AIPWE), which requires models for both $\mu(A, \boldsymbol{X})$ and $P(A = 1|\boldsymbol{X})$. The AIPWE results is a gain in efficiency relative to IPWE and has a double robustness property. In a simulation study, the RG method was the best if the model for $\mu(A, \boldsymbol{X})$ was correctly specified, but was not robust to misspecification of $\mu(A, \boldsymbol{X})$. With correct specification of $\mu(A, \boldsymbol{X})$, the AIPWE method was not quite as efficient as RG.

For the misspecified model for $\mu(A, \boldsymbol{X})$, residual plots would immediately recognize the model as providing a poor fit to the data. In this article we examine the relative merits of RG, IPWE and AIPWE when one uses a model for $\mu(A, \boldsymbol{X})$ which better fits the data.

2. Review of Methods

Let $Y(g)$ be the response for a patient who follows regime g. For a randomly chosen patient from a population the expected response if regime $g(\boldsymbol{X})$ is followed is given by $E(Y(g)) = E_X\big[\mu(1, \boldsymbol{X})g(\boldsymbol{X}) + \mu(0, \boldsymbol{X})\{1 - g(\boldsymbol{X})\}\big]$. The optimal treatment regime is $g^{\text{opt}}(\boldsymbol{X}) = I\{\mu(1, \boldsymbol{X}) > \mu(0, \boldsymbol{X})\}$. Let $\hat{g}$ denote an estimated regime that is derived from a dataset.

Denote by $Q(g)$ the average value of the expected response for subjects in a future population of very large size N if regime g were to be used, where $Q(g)$ is given by

$$Q(g) = \frac{1}{N}\sum_{i=1}^{N}\big[\mu(1, \boldsymbol{X}_i)g(\boldsymbol{X}_i) + \mu(0, \boldsymbol{X}_i)\{1 - g(\boldsymbol{X}_i)\}\big]. \quad (1)$$

Larger values of $Q(g)$ are better. Thus, when $\mu(A, \boldsymbol{X})$ is known, the success of different methods for estimating g can be based on $Q(\hat{g})$ and also compared to $Q(g^{\text{opt}})$.

2.1. *Regression Method*

The RG method is to posit a parametric regression model for $\mu(A, \boldsymbol{X}) = \mu(A, \boldsymbol{X}; \boldsymbol{\beta})$, estimate $\boldsymbol{\beta}$ from the data, then $\widehat{g}^{\text{opt}}_{\text{reg}}(\boldsymbol{X}) = I\{\mu(1, \boldsymbol{X}; \widehat{\boldsymbol{\beta}}) > \mu(0, \boldsymbol{X}; \widehat{\boldsymbol{\beta}})\}$. Below we will also consider alternative nonparametric regression models for $\mu(A, \boldsymbol{X})$.

2.2. *Inverse Probability Weighted Estimators*

For the IPWE method, a parametric form for $g(\boldsymbol{X}) = g(\boldsymbol{X}; \boldsymbol{\eta})$ is specified. For fixed $\boldsymbol{\eta}$, define $C_{\boldsymbol{\eta},i} = A_i g(\boldsymbol{X}_i; \boldsymbol{\eta}) + (1 - A_i)\{1 - g(\boldsymbol{X}_i; \boldsymbol{\eta})\}$ and $\pi(\boldsymbol{X}) = P(A = 1|\boldsymbol{X})$. Then the expected population average outcome is $\frac{1}{n}\sum_{i=1}^{n} C_{\eta,i}Y_i/\pi_C(\boldsymbol{X}_i)$ which is maximized over $\boldsymbol{\eta}$ to give $\widehat{g}^{\text{opt}}_{\text{IPW}}(\boldsymbol{X}) = g(\boldsymbol{X}; \widehat{\boldsymbol{\eta}})$ where $\pi_C(\boldsymbol{X}_i) = \pi(\boldsymbol{X}_i)^{A_i}\{1 - \pi(\boldsymbol{X}_i)\}^{1-A_i}$. For a randomized study the propensity score $\pi(\boldsymbol{X})$ is estimated by the sample proportion assigned to treatment 1, which will be close to 0.5. For a non-randomized study logistic regression is used to estimate $\pi(\boldsymbol{X})$.

For the AIPWE method, $\boldsymbol{\eta}$ is obtained by maximizing

$$\text{AIPWE}(\eta) = \frac{1}{n}\sum_{i=1}^{n} \frac{C_{\boldsymbol{\eta},i} Y_i}{\pi_C(\boldsymbol{X}_i)} - \frac{C_{\boldsymbol{\eta},i} - \pi_C(\boldsymbol{X}_i)}{\pi_C(\boldsymbol{X}_i)} m(X_i; \eta, \widehat{\boldsymbol{\beta}}) \tag{2}$$

over $\boldsymbol{\eta}$, where $m(\boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\beta}) = \mu(1, \boldsymbol{X}; \boldsymbol{\beta}) g(\boldsymbol{X}; \boldsymbol{\eta}) + \mu(0, \boldsymbol{X}; \boldsymbol{\beta})\{1 - g(\boldsymbol{X}; \boldsymbol{\eta})\}$.

Zhang et al. (2012) also considered the consistency properties and calculation of standard errors for $\widehat{\boldsymbol{\eta}}$, we will not consider these in the current article.

3. Simulation Study

In the simulation study in Zhang et al. (2012), in Table 8 of the Supplementary Materials, data were generated from a true model $Y_i = \mu(A_i, \boldsymbol{X}_i) + e_i$, where $e_i \sim N(0, 1)$ and

$$\mu(A, \boldsymbol{X}) = \exp\{2.0 - 1.5X_1^2 - 1.5X_2^2 + 3.0X_1X_2 + A(-0.1 - X_1 + X_2 + 0.2X_3)\},$$

where X_{i1} and X_{i2} were $U(-1.5, 1.5)$ and X_{i3} and A_i were Bern(0.5). For this model $g^{\text{opt}}(\boldsymbol{X}) = I(-0.1 - X_1 + X_2 + 0.2X_3 > 0)$. They considered two parametric regression models for $\mu(A, \boldsymbol{X})$, a correctly specified model of the form

$$\mu_t(A, \boldsymbol{X}; \boldsymbol{\beta}) = \exp\{\beta_0 + \beta_1 X_1^2 + \beta_2 X_2^2 + \beta_3 X_1 X_2 + A(\beta_4 + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3)\} \tag{3}$$

and a misspecified simple linear model of the form

$$\mu_{\text{msl}}(A, \boldsymbol{X}; \boldsymbol{\beta}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + A(\beta_4 + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3). \tag{4}$$

From standard residual plots it is obvious that the misspecified model gives a very bad fit to the data, and would not be seriously entertained, particularly for the RG method. Inspection of the data suggests that some transformation of the response Y may lead to an improved fit. Although $\log(Y)$ might appear to be a natural choice, it is not possible because a small fraction of the Y's are negative, thus we choose $Y^{1/3}$ as an approximation. Thus the question is, if one used a better fitting model for Y in both the RG and AIPWE methods would the results improve? We consider two parametric models, as well as a non-parametric estimator. The first misspecified parametric model recognizes the benefit of a transformation, and the second also recognizes the need for quadratic terms and interactions. In these models we develop predictions for $Z = Y^{1/3}$, and then cube these predictions of Z to obtain predictions of Y. The simple misspecified cube root model is given by

$$E(Z) = \mu_{\text{ms33}}(A, \boldsymbol{X}; \boldsymbol{\beta}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + A(\beta_4 + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3). \tag{5}$$

and the misspecified complex cube root model is given by

$$\mu_{mc33}(A, \boldsymbol{X}; \boldsymbol{\beta}) = \mu_{\text{ms33}}(A, \boldsymbol{X}; \boldsymbol{\beta}) + \sum_{j=1}^{2} \beta_{j+7} X_j^2 + \beta_{10} X_1 \times X_2 + \beta_{11} X_1 \times X_3 + \beta_{12} X_2 \times X_3. \tag{6}$$

Standard model assessment methods would recognize some lack of fit for μ_{mc33}, although it is a noticeable improvement over μ_{msl} and μ_{ms33}.

In the other approach, we used random forests as a non-parametric estimator of $\mu(A, \boldsymbol{X})$, and denote the estimate by $\widehat{\mu}(A, \boldsymbol{X})$. The RG_{rf} method consists of maximizing

$$\frac{1}{n}\sum_{i=1}^{n} \left[\widehat{\mu}(1, \boldsymbol{X}_i) g(\boldsymbol{X}_i; \boldsymbol{\eta}) + \widehat{\mu}(0, \boldsymbol{X}_i)\{1 - g(\boldsymbol{X}_i; \boldsymbol{\eta})\}\right] \tag{7}$$

with respect to $\boldsymbol{\eta}$. While we present results for random forests, other non-parametric estimators could be considered. To implement random forests, with $Y^{1/3}$ as the response, we used the function *randomforest* in R, using default settings except that the number of trees was 1000. Similar to previous work (Foster et al., 2011), we found that the performance of random forests was improved by using $A, X_k, X_k^2, X_k I(A = 1)$ and $X_k I(A = 0)$ for $k = 1$ to p as input covariates. We note that random forests with $Y^{1/3}$ as the response gave a very mild improvement over random forests with Y as the response.

To fit the linear model in equations (4)–(6), the R function *lm* was used. To fit the non-linear model in equation (3), the R function *nlsLM* was used. To maximize the criteria in equations (2) and (7), we used the R function *genoud*, as described in Zhang et al. (2012).

3.1. *Results for Three Covariates*

In our simulation study 1000 datasets, each of size 500 was generated. We report in Table 1 two quantities: the average of the ratio $Q(\widehat{g})/Q(g^{\text{opt}})$ and the average fraction who would be treated if, following each trial, $\widehat{g}$ were to be used. For each of the 1000 datasets $Q(\widehat{g})$ and $Q(g^{\text{opt}})$ were calculated from equation (1), with N=1,000,000.

The second and third columns show the results for three covariate, labeled as Case A, matching the situation considered in (Zhang et al., 2012). For this setting $Q(g = 0) = 3.02$, $Q(g = 1) = 3.48$, and $Q(g^{\text{opt}}) = 3.95$. The first two rows are for ideal, but not applicable in practice, methods in which the structure of the true model for $\mu(A, \boldsymbol{X})$ is known. Amongst these two, RG_t is slightly better than AIPWE_t, achieving the desired values of 1 and 0.5 for the Ratio to optimal and the Fraction treated, respectively. Amongst the applicable methods, RG_{mc33} generally improves on RG_{msl} and RG_{ms33}, and RG_{rf} is the much better than both. Amongst the AIPWE methods they are all preferable to IPWE, with AIPWE_{rf} being the best.

Also of note is that the inverse probability methods tend to recommend treating closer to the true 50% fraction of patients, than the regression methods. The regression methods tend to include too many subjects in the region $\widehat{g}(\boldsymbol{X}) = 1$.

In the fourth and fifth columns of the table we show results for Case B, a situation in which the optimal regime is that 90% of the subjects should be treated. The data were generated from the model $\mu(A, \boldsymbol{X}) = \exp\{2.0 - 1.5X_1^2 - 1.5X_2^2 + 3.0X_1X_2 + A(-0.1 - X_1 + X_2 + 0.2X_3)\}$, where X_{i1} was $U(-1.5, 1.5)$, X_{i2} was $U(0.2, 3)$, X_{i3} was Bern(0.6) and A_i was Bern(0.5). For this setting $Q(g = 0) = 1.66$, $Q(g = 1) = 2.51$, and $Q(g^{\text{opt}}) = 2.64$.

Table 1

Simulation results, randomized studies. Ratio to optimal is $Q(\widehat{g})/Q(g^{\text{opt}})$. Fraction treated denotes the fraction that would be treated in a future population if regime $\widehat{g}$ was followed. Case A: 3 independent covariates. Optimal fraction treated = 0.5. Case B: 3 independent covariates. Optimal fraction treated = 0.902. Case C: 3 independent covariates. Optimal g includes interaction. Optimal fraction treated = 0.5. Case D: 15 independent covariates. Optimal fraction treated = 0.500. The subscripts denote the model that was used for estimating $\mu(A, \boldsymbol{X})$, t=true, msl=misspecified simple linear, ms33=misspecified simple with $Y^{1/3}$, mc33=misspecified complex with $Y^{1/3}$, rf=random forest.

	Case A p=3, indep		Case B p=3, indep		Case C p=3, interaction		Case D p=15, indep	
Method	Ratio to optimal	Fraction treated	Ratio to optimal	Fraction treated	Ratio to optimal	Fraction treated	Ratio to optimal	Fraction treated
Assuming form of true model is known								
RG_t	1.000	0.50	0.999	0.90	1.000	0.50	1.000	0.50
AIPWE_t	0.997	0.50	0.995	0.90	0.996	0.51	0.997	0.50
Form of model is unknown								
RG_{msl}	0.925	0.67	0.946	0.93	0.922	0.63	0.893	0.62
RG_{ms33}	0.927	0.58	0.936	0.91	0.924	0.56	0.883	0.55
RG_{mc33}	0.948	0.58	0.936	0.91	0.942	0.57		
RG_{rf}	0.990	0.53	0.990	0.92	0.977	0.53	0.904	0.68
IPWE	0.971	0.49	0.970	0.89	0.956	0.50	0.879	0.58
$\text{AIPWE}_{\text{msl}}$	0.984	0.50	0.977	0.87	0.965	0.52	0.884	0.62
$\text{AIPWE}_{\text{ms33}}$	0.978	0.49	0.970	0.90	0.963	0.50	0.885	0.59
$\text{AIPWE}_{\text{mc33}}$	0.989	0.49	0.966	0.91	0.976	0.50		
AIPWE_{rf}	0.990	0.50	0.992	0.89	0.976	0.50	0.896	0.60

We considered four parametric outcome regression models. The first one is $\mu_t(A, \boldsymbol{X}; \boldsymbol{\beta})$, the correct nonlinear regression model, as given in equation (3); the other three were the misspecified models $\mu_{\text{msl}}(A, \boldsymbol{X}; \boldsymbol{\beta})$, $\mu_{\text{ms33}}(A, \boldsymbol{X}; \boldsymbol{\beta})$, and $\mu_{\text{mc33}}(A, \boldsymbol{X}; \boldsymbol{\beta})$, as given in equations (4)–(6). The results again show the benefit of using random forests to estimate $\mu(A, \boldsymbol{X})$ in both RG and AIPWE methods and that again RG_{rf} has similar performance as AIPWE_{rf}.

We also considered a case similar to case A, but in which the covariates were correlated. The results were very similar to the uncorrrelated case and are not presented here.

In the sixth and seventh columns of the table we show results for Case C, a situation where the optimal g is not determined by a linear combination of the covariates. The data were generated from the model $\mu(A, \boldsymbol{X}) = \exp\{2.0 - 1.5X_1^2 - 1.5X_2^2 + 3.0X_1X_2 + A(-0.1 - X_1 + X_2 + 0.2X_3 + 0.5X_1X_3)\}$, where X_{i1} and X_{i2} were $U(-1.5, 1.5)$, X_{i3} and A_i were Bernoulli(0.5). For this setting $Q(g = 0) = 3.02$, $Q(g = 1) = 3.49$, and $Q(g^{\text{opt}}) = 3.99$. The optimal $g(\boldsymbol{X})$ is $I(-0.1 - X_1 + X_2 + 0.2X_3 + 0.5X_1X_3 > 0)$.

We considered four parametric outcome regression models. The first one is $\mu_t(A, \boldsymbol{X}; \boldsymbol{\beta})$

$$\mu_t(A, \boldsymbol{X}; \boldsymbol{\beta}) = \exp\{\beta_0 + \beta_1 X_1^2 + \beta_2 X_2^2 + \beta_3 X_1 X_2 + A(\beta_4 + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3 + \beta_8 X_1 X_3)\} \quad (8)$$

corresponding to the correct nonlinear regression model; the other three were $\mu_{\text{msl}}(A, \boldsymbol{X}; \boldsymbol{\beta})$ and $\mu_{\text{ms33}}(A, \boldsymbol{X}; \boldsymbol{\beta})$, and $\mu_{\text{mc33}}(A, \boldsymbol{X}; \boldsymbol{\beta})$ as given in (4)–(6).

The results are similar to those for Case A, with for the RG methods a mild improvement by using the complex parametric model and substantial improvement by using random forests. The results again show the benefit of using random forests to estimate $\mu(A, \boldsymbol{X})$ in the AIPWE methods. The fact that the optimal g is not within the class of models being estimated does not seem to have negatively impacted the performance of the methods.

3.2. *Results for 15 Covariates*

The above results are for a small number of three covariates. With a larger number of covariates, the task of building models for $\mu(A, \boldsymbol{X})$ is more challenging. Fitting parametric models with many quadratic terms and interactions is not feasible, or would require variable selection. The ability of non-parametric regression methods, such as random forests, to give reliable predictions decreases with increasing p. The performance of the AIPWE methods is also likely to deteriorate with larger p, because the maximization in equation (2) will give poorer estimates of $\boldsymbol{\eta}$. To investigate this, we considered a situation of 15 covariates, where the true model for $\mu(A, \boldsymbol{X})$ was

$$\mu(A, \boldsymbol{X}) = \exp\{2.0 - 1.5X_1^2 - 1.5X_2^2 + 3.0X_1X_2 + A(-0.1 - X_1 + X_2 + 0.2X_3)\},$$

with corresponding $g^{\text{opt}}(\boldsymbol{X}) = I(X_2 > X_1 - 0.2X_3 + 0.1)$, and where X_1 and $X_2 \sim U(-1.5, 1.5)$, $X_3 \sim$ Bern(0.5), $X_4, X_5, X_7, X_8, X_{10}, X_{11}, X_{13}, X_{14} \sim U(-1.5, 1.5)$ and $X_6, X_9, X_{12}, X_{15} \sim$ Bern(0.5). For this setting $Q(g = 0) = 3.02$,

Table 2
Simulation results, nonrandomized studies. Ratio to optimal is $Q(\widehat{g})/Q(g^{\text{opt}})$. Fraction Treated denotes the fraction that would be treated in a future population if regime $\widehat{g}$ was followed. Two independent covariates. Optimal fraction treated = 0.5. The subscripts denote the model that was used for estimating $\mu(A, \boldsymbol{X})$, t=true, msl=misspecified simple linear, ms33=misspecified simple with $Y^{1/3}$, mc33=misspecified complex with $Y^{1/3}$, rf=random forest.

Method	Ratio to optimal	Fraction treated
$\text{RG}\mu_{\text{t}}$	1.000	0.47
$\text{RG}\mu_{\text{msl}}$	0.878	0.25
$\text{RG}\mu_{\text{ms33}}$	0.861	0.18
$\text{RG}\mu_{\text{mc33}}$	0.936	0.50
RG_{rf}	0.994	0.49
Propensity score model correct		
IPWE	0.979	0.47
$\text{AIPWE}\mu_{\text{t}}$	0.998	0.47
$\text{AIPWE}\mu_{\text{msl}}$	0.988	0.47
$\text{AIPWE}\mu_{\text{ms33}}$	0.984	0.47
$\text{AIPWE}\mu_{\text{mc33}}$	0.991	0.46
AIPWE_{rf}	0.995	0.47
Propensity score model incorrect		
IPWE	0.921	0.33
$\text{AIPWE}\mu_{\text{t}}$	0.998	0.47
$\text{AIPWE}\mu_{\text{msl}}$	0.961	0.39
$\text{AIPWE}\mu_{\text{ms33}}$	0.934	0.35
$\text{AIPWE}\mu_{\text{mc33}}$	0.982	0.42
AIPWE_{rf}	0.995	0.47

$Q(g = 1) = 3.48$, and $Q(g^{\text{opt}}) = 3.95$. The linear combination that determines the estimated g could include 15 variables.

We considered three possible parametric outcome regression models. The first one was $\mu_t(A, \boldsymbol{X}; \boldsymbol{\beta}) = \exp\{\beta_0 + \beta_1 X_1^2 + \beta_2 X_2^2 + \beta_3 X_1 X_2 + A(\beta_4 + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3)\}$, which corresponds to the correct nonlinear regression model; the second one is $\mu_{\text{msl}}(A, \boldsymbol{X}; \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^{15} \beta_j X_j + A(\beta_{16} + \sum_{j=1}^{15} \beta_{j+16} X_j)$. The third one is $\mu_{\text{ms33}}(A, \boldsymbol{X}; \boldsymbol{\beta})$, which is the same as $\mu_{\text{msl}}(A, \boldsymbol{X}; \boldsymbol{\beta})$ except that the response is $Y^{1/3}$. Fitting $\mu_{\text{mc33}}(A, \boldsymbol{X}; \boldsymbol{\beta})$ was not feasible in this case. The RG_{rf} and AIPWE methods were also implemented.

The results for Case D, given in the eighth and ninth columns, differ from those of Case A. Here the RG method with a simple misspecified linear model has properties as good as those from AIPWE using this misspecified model and better than the IPWE method. Again we see that both RG and AIPWE methods are improved by the use of random forests. The general performance of all the methods, is clearly worse when there are more covariates.

3.3. *Results for Non-Randomized Trial Setting*

For this situation the RG methods are unchanged, but the IPWE and AIPWE require formulating and fitting an additional model for $P(A = 1|\boldsymbol{X})$.

In the first simulation scenario presented by Zhang et al. (2012), they generated data from a true model of the form $Y_i = \mu(A_i, \boldsymbol{X}_i) + \epsilon_i$, where $\epsilon_i \sim N(0, 1)$ and

$$\mu(A, \boldsymbol{X}) = \exp\{2.0 - 1.5X_1^2 - 1.5X_2^2 + 3.0X_1X_2 + A(-0.1 - X_1 + X_2)\}, \quad (9)$$

where X_1 and X_2 were independent $U(-1.5, 1.5)$. The treatment group indicator A_i was determined by the model $\text{logit}\{P(A = 1|\boldsymbol{X})\} = -0.1 + 0.8X_1^2 + 0.8X_2^2$.

For model (9) $g^{\text{opt}}(\boldsymbol{X}) = I(-0.1 - X_1 + X_2 > 0)$, and $E\{Y(g^{\text{opt}})\} = 3.71$. Two regression models for $\mu(A, \boldsymbol{X})$ were considered, a correctly specified model of the form

$$\mu_t(A, \boldsymbol{X}; \boldsymbol{\beta}) = \exp\{\beta_0 + \beta_1 X_1^2 + \beta_2 X_2^2 + \beta_3 X_1 X_2 + A(\beta_4 + \beta_5 X_1 + \beta_6 X_2)\}$$

and a misspecified simple linear model of the form

$$\mu_{\text{msl}}(A, \boldsymbol{X}; \boldsymbol{\beta}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2).$$

The misspecified model μ_{ms33} is $E(Z) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)$ and misspecified model μ_{mc33} is

$$E(Z) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2) + \sum_{j=1}^{2} \beta_{j+5} X_j^2 + \beta_8 X_1 \times X_2,$$

where $Z = Y^{1/3}$. Method rf applies the random forest approach. The input covariates used were A, X_k, X_k^2, $X_k I(A = 1)$ and $X_k I(A = 0)$ for $k = 1$ to 2, and the response variable is Z.

The propensity score model for $P(A = 1|\boldsymbol{X})$, required for IPWE and AIPWE, is either correctly specified as $\text{logit}\{P(A = 1|\boldsymbol{X})\} = \gamma_0 + \gamma_1 X_1^2 + \gamma_2 X_2^2$ or incorrectly specified as $\text{logit}\{P(A = 1|\boldsymbol{X})\} = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2$.

The results show that the use of random forests in the RG methods is as good as any other method, and that poorly fitting parametric models for both RG and for AIPWE when the propensity score model is incorrect can lead to regimes with noticeably worse properties.

4. Discussion

Zhang et al. (2012) article illustrates that regression methods may not be robust to model misspecification, and that AIPWE methods do have an appealing robustness property. However, this robustness property should not be an excuse for not seeking reasonably fitting models for the data. We demonstrate in a small simulation study, that modeling the

response for the regression method with a better fitting parametric model leads to some improvement, while using a readily available non-parametric method removes concerns about non-robustness of the regression method. Furthermore, the properties of both regression and augmented inverse probability weighted methods are improved by using the non-parametric method for the response compared to parametric models, and are quite similar. Thus the extra modeling needed for the AIPWE is not doing any harm, but also may not be necessary.

Acknowledgements

This work was partially funded by US National Institutes of Health grant CA 129102 and by Eli Lilly.

References

Foster, J. C., Taylor, J. M. G., and Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine* **30**, 2867–2880.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **168**, 1010–1018.

Received December 2013. *Revised June* 2014.
Accepted July 2014.

The authors replied as follows:

We applaud Taylor, Cheng, and Foster (henceforth TCF) for carrying out additional empirical studies of methods for estimating optimal treatment regimes, as further elucidation of the relative performance of competing methods is sorely needed. We hope that the evidence suggesting that these methods can perform well and yield comparable results under conditions likely to hold in practice will encourage more widespread interest in estimation of optimal treatment regimes.

TCF consider the situation where the class of regimes $\mathcal{G}_\eta$ of interest has elements of the form $g(\boldsymbol{X};\boldsymbol{\eta}) = I(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X} > 0)$. They study the estimators *IPWE* and *AIPWE* for an optimal regime proposed in our 2012 paper (Zhang et al., 2012b), which are based on maximizing in $\boldsymbol{\eta}$ inverse probability weighted estimators for the expected outcome, or *value*, under a regime in a specified class $\mathcal{G}_\eta$; we have referred to estimators for an optimal regime found by maximizing an estimator for the value in $\boldsymbol{\eta}$ as *value search* or policy search estimators. These are compared to two competing approaches. The first is the regression estimator *RG*, which is based on a posited parametric model $\mu(a,\boldsymbol{x};\boldsymbol{\beta})$ for $\mu(a,\boldsymbol{x}) = E(Y|A=a,\boldsymbol{X}=\boldsymbol{x})$ that induces the class of regimes $\mathcal{G}_\eta$ when the model involves an interaction term of the form $a(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x})$. The estimated optimal regime is found directly as $I\{\mu(1,\boldsymbol{x};\widehat{\boldsymbol{\beta}}) > \mu(0,\boldsymbol{x};\widehat{\boldsymbol{\beta}})\}$.

In the approach TCF call RG_{rf}, a different estimator for the value from those in *IPWE* or *AIPWE*, given in their Equation (6), is maximized in $\boldsymbol{\eta}$. This method is proposed by us in Zhang et al. (2012a), in which we expressed (6) equivalently in terms an estimator for the *contrast function* $C(\boldsymbol{x}) = \mu(1,\boldsymbol{x}) - \mu(0,\boldsymbol{x})$. In particular, we suggested estimating an optimal regime of a specified form $g(\boldsymbol{X};\boldsymbol{\eta})$ by maximizing in $\boldsymbol{\eta}$ the estimator for the value given by $n^{-1}\sum_{i=1}^{n} g(\boldsymbol{X}_i;\boldsymbol{\eta})\widehat{C}(\boldsymbol{X}_i)$, where $\widehat{C}(\boldsymbol{x})$ is formed by representing $\mu(a,\boldsymbol{x})$ in $C(\boldsymbol{x})$ by a nonparametric estimator $\widehat{\mu}(a,\boldsymbol{x})$, such as support vector regression (Vapnik, Golowich, and Smola, 1997) or boosting (Freund and Schapire, 1997). TCF use random forests, noting that any flexible nonparametric estimator could be used. We thus disagree with TCF's characterization of this as a "regression" method and find reference to it as "*RG* with random forests" to be a bit misleading. This method, like *IPWE* and *AIPWE*, is a value search approach. These three methods are thus different in spirit and construction from *RG*, which bases the estimated optimal regime directly on a fitted regression model. We did not evaluate the performance of the contrast-based value search approach empirically in Zhang et al. (2012a), so the studies by TCF in which it is implemented using the particular choice of random forests, RG_{rf}, fill an important gap and demonstrate its feasibility and robustness.

For *RG*, TCF rightly consider misspecified parametric models that are closer to the true $E(Y|A,\boldsymbol{X})$ than the linear model we adopted to represent model misspecification in Zhang et al. (2012b), as would likely be formulated by a careful data analyst. We fully agree with their contention that *RG* can perform well under these conditions, as the simulations they present demonstrate. TCF also confirm our finding that *IPWE* is inferior to the other methods. The evidence from their studies along with that in our paper demonstrates that all of *RG*, *AIPWE*, and RG_{rf}, and especially the latter two value search approaches, lead to high quality estimated regimes, providing the data analyst with a range of options. In fact, the analyst can employ competing approaches and compare the results to gain an understanding of sensitivity to modeling choices.

A limitation of *RG* is that, as noted above, the class of regimes considered and the resulting estimated regimes are dictated by the form of the posited parametric regression model. On the other hand, if one were to use flexible nonparametric estimators like random forests to represent $\mu(a,\boldsymbol{x})$ directly in the regression-based method, which seems to us an approach for which "RG_{rf}" is a more appropriate acronym than its use by TCF, the result would be estimated regimes of a "black box" nature, which may elicit skepticism from clinicians. In contrast, the value search approaches, *IPWE*, *AIPWE*, and TCF's RG_{rf}, search within a user-defined class of regimes whose specification need not be connected with the form of models for $\mu(a,\boldsymbol{x})$. This is advantageous if interest is in regimes having a specific form on the basis of interpretability, cost, or feasibility in practice that may not be induced straightforwardly from models for $\mu(a,\boldsymbol{x})$. For instance, as we demonstrated in Zhang et al. (2012b), the class of regimes can be restricted to have elements defined by rectangular regions; for example, $\boldsymbol{X} = (X_1, X_2)^{\mathrm{T}}$, $g(\boldsymbol{X};\boldsymbol{\eta}) = I(X_1 > \eta_1, X_2 > \eta_2)$, which clinicians may find more interpretable than regimes involving linear combinations of covariates.

TCF consider primarily the case of randomized studies, where the propensity scores $\pi(\boldsymbol{x})$ are known and generally

constant in $\boldsymbol{x}$. Here, *IPWE* and *AIPWE* are based on estimators for the value of a regime in $\mathcal{G}_\eta$ that are guaranteed by construction to be consistent, which, intuitively, would be expected to lead to well-performing estimated optimal regimes. Moreover, these methods require no additional modeling, as the propensity score is estimated by the sample randomization proportion. The estimator for the value in (6) of TCF that forms the basis for RG_{rf} is, in contrast, not consistent unless the model for $\mu(a, \boldsymbol{x})$ is correctly specified, and the *RG* method depends critically on a correct model. As TCF demonstrate, this may be of little consequence with RG_{rf} and a sufficiently flexible representation for $\mu(a, \boldsymbol{x})$ or with *RG* and a "nearly correct" parametric model, although the evidence in TCF is less compelling for the latter estimator. Overall, we agree with TCF that the value search estimators *AIPWE*, incorporating a flexible model for $\mu(a, \boldsymbol{x})$ in the "augmentation term," and RG_{rf} are the most promising in this setting. From a theoretical point of view, an advantage of *AIPWE* is that in this setting it yields the locally efficient estimator for the value; see Robins and Ritov (1997).

In an observational study, *AIPWE* is based on a value estimator that is doubly robust; that is, guaranteed to be consistent as long as at least one of the propensity score model or model for $\mu(a, \boldsymbol{x})$ is correctly specified, whereas, as TCF note, RG_{rf} is not doubly robust. We agree with TCF that, if one has considerable confidence in the nonparametric random forest representation for the contrast function, including its incorporated adjustment for confounding, the additional protection afforded by the *AIPWE* may be unnecessary. However, *AIPWE* implemented with careful modeling of the propensity score in the same spirit as TCF propose in *RG* could provide the analyst with additional trust in the robustness of results.

A challenge with all of the value search methods *IPWE*, *AIPWE*, and RG_{rf} is that the maximization of the value estimator in $\boldsymbol{\eta}$ is a nonsmooth optimization problem that cannot be addressed using standard optimization methods. In problems where the restricted class of regimes involves rich covariate information, so that $\boldsymbol{\eta}$ is high-dimensional, implementation becomes computationally prohibitive and the quality of estimation will be degraded.

One practical approach to circumventing this difficulty is described in Zhang et al. (2012a), where we demonstrated how the problem of maximizing value search estimators in $\boldsymbol{\eta}$ can be recast as minimizing a weighted classification error; see also Zhao et al. (2012). Thus, estimation of an optimal treatment regime can be likened to a classification problem, viewing $g(\boldsymbol{x}; \boldsymbol{\eta})$ as a classifier, with the class of regimes $\mathcal{G}_\eta$ determined by the choice of classifier; for example, classification and regression trees (Breiman et al., 1984) or support vector machines (Cortes and Vapnik, 1995). In this formulation, the "class label" and "weight" are functions of the estimated contrast function. Existing software for carrying out the minimization for a given choice of classifier can then be used to estimate an optimal regime in this class. Although this is also a nonstandard optimization problem, an advantage in practice is that computational techniques to approximate it efficiently and to carry out the variable selection involved are embedded in off-the-shelf software.

A possible advantage of *AIPWE* over *RG* is the extension to more than one treatment decision point. The extension of *RG*, Q-learning, requires positing a sequence of regression models at each decision point that ideally must be compatible with one another. In practice, such a specification is almost impossible (but see Laber, Linn, and Stefanski, in press), so that the models at decision points other than the last one are almost certainly misspecified, even if flexible methods are used, which will compromise the quality of estimated regimes. The extension of *AIPWE* we present in Zhang et al. (2013) ideally requires specification of compatible such models, but only for the purpose of gaining efficiency and ensuring approximate double robustness. Extension of TCF's RG_{rf} and related contrast-based value search estimators to this setting should be investigated. More generally, further research is needed to clarify the performance of approaches in the multiple decision setting.

Given the well-performing options available for estimating optimal regimes, we believe that the most pressing challenge is that the methodological advances have far outpaced current practice. We must encourage our clinician collaborators and practicing biostatisticians to consider estimation of dynamic treatment regimes as a meaningful, primary data-analytic objective. Although this perspective has been embraced by some researchers in the behavioral sciences, it is not as prevalent in chronic disease research, where interest focuses primarily on identifying subgroups of patients to whom treatment may be targeted; that is, identifying the "right patient for the treatment." Thinking in terms of optimal treatment regimes, so identifying "the right treatment for the patient," offers a valuable complementary perspective. The critical next step for the treatment regime research community is a proactive effort to communicate the concepts and methods and their scientific relevance to health sciences researchers more broadly.

That said, an outstanding methodological challenge is inference for the estimated regime. The value of an estimated regime is equivalent to the weighted test error of an estimated classifier and is thus a data-dependent, nonsmooth functional of the underlying generative distribution (Laber and Murphy, 2011; Chakraborty, Laber, and Zhao, 2014). Standard asymptotic methods for inference, including the bootstrap and series approximations, do not apply without modification, and the small sample performance of these methods can be quite poor under some generative models. Inference for the parameters indexing the optimal regime has been another focus for inference (Robins, 2004; Chakraborty, Murphy, and Strecher, 2010; Laber et al., 2014). However, it is not clear that this is an appropriate target for inference for value search estimators where the objective is to estimate a high-quality regime within a prespecified class, which need not be assumed to contain the true optimal regime.

We thank TCF again for a thoughtful and important demonstration of the relative merits of estimators for optimal dynamic treatment regimes. Their findings, in conjunction with other work cited herein, make a strong case for the use of value search estimators in practice.

Acknowledgements

This work was supported by NIH grants P01 CA142538, R01 HL118336, and R01 CA085848.

References

Breiman, L, Freidman, J. L., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, California: Wadsworth.

Chakraborty, B., Laber, E. B., and Zhao, Y. Q. (2014). Inference about the expected performance of a data-driven dynamic treatment regime. *Clinical Trials* **11**, 408–417.

Chakraborty, B., Murphy, S., and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research* **19**, 317–343.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning* **20**, 273–297.

Freund, B. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* **55**, 119–139.

Laber, E. B., Linn, K. A., and Stefanski, L. A. (in press). Interactive model building for Q-learning. *Biometrika*.

Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E., and Murphy, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electronic Journal of Statistics* **8**, 1225–1272.

Laber, E. B. and Murphy, S. A. (2011). Adaptive confidence intervals for the test error in classification. *Journal of the American Statistical Association* **106**, 904–913.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*, 189–326. New York: Springer.

Robins, J. M. and Ritov, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semiparametric models. *Statistics in Medicine* **16**, 285–319.

Vapnik, V., Golowich, S., and Smola, A. (1997). Support vector regression for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems* **9**, 281–287.

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., and Laber, E. B. (2012a). Estimating optimal treatment regimes from a classification perspective. *Stat* **1**, 103–114.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012b). A robust method for estimating optimal treatment regimes. *Biometrics* **68**, 1010–1018.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100**, 681–694.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* **107**, 1106–1118.

Baqun Zhang
School of Statistics
Renmin University of China
Beijing 100872, China
email: zhangbaqun@ruc.edu.cn

Anastasios A. Tsiatis
Department of Statistics,
Box 8203,
North Carolina State University,
Raleigh, North Carolina 27695, U.S.A.
email: tsiatis@ncsu.edu

Eric B. Laber
Department of Statistics,
Box 8203,
North Carolina State University,
Raleigh, North Carolina 27695, U.S.A.
email: eblaber@ncsu.edu

and

Marie Davidian
Department of Statistics,
Box 8203,
North Carolina State University,
Raleigh, North Carolina 27695, U.S.A.
email: davidian@ncsu.edu